

WO 03/077579

1

Radio resource allocation in a radio communication network**Technical field**

The invention relates to a method for allocating radio resources of a radio communication network to a plurality of users, where a user is allocated a certain transmission capacity.

- 5 The invention further relates to a radio network as well as a device with means for an allocation of radio resources to a plurality of users, where a user is allocated a certain transmission capacity.

Prior art

In modern telecommunication networks, transmission frequencies and transmission time typically are limited. That is why the availability of radio resources is one of the most valuable factors in communication systems, particularly in radio communication networks.

5 Many different schemes, e.g. time, frequency or code division multiple access systems (TDMA, FDMA, CDMA respectively), have been developed to increase the available resources for a given transmission system. While these transmission schemes work very well, the efficient allocation of the available resources to different users is a challenge. The goal is to allocate resources very quickly to a user who has information to transmit and to immediately deallocate it from the user, when he has nothing to transmit in order to allocate it to another user who has something to transmit.

10

In resource allocation, it is not only important, how much of the available resource, i.e. data rate, is allocated to a user, but also how fast it is allocated to him. Radio resource allocation is particularly difficult, when a user is sending information intermittently. Ideally, the 15 necessary resources should be allocated to the user as soon as he asks for it and then the resources should be taken away, again without delay, when he has nothing more to send.

Existing radio networks try to solve this problem in different ways. Some systems start with allocating several short blocks of a given resource after the other and then increase the duration of the blocks until the user is allocated a continuous amount of resource.

20 Others allocate a certain resource to a particular user for a long time right from the start.

In the first case, the user can not send data continuously from the beginning, but only during the time when the resource is available. Hence he sometimes has to wait for allocation even if he has data to send. This results in a lowered throughput for this user, but since the resource can be shared with other users, less resources are wasted. In the second 25 case, the user will have a high throughput rate if he has data to send. However, if he is only

able to send data intermittently, some of the resources allocated to the user are wasted, because he was not using it all the time.

Each allocation method has its advantages and disadvantages. One method is best suited for a first application while another method is best suited for another application. The problem is, that the network provider does not know, how the actual allocation method performs. He gets no corresponding feedback.

Summary of the invention

It is therefore an object of the invention to provide a method of allocating radio resources of a radio communication network to a plurality of users, permitting an efficient radio resource allocation.

The object of the invention is achieved by the method defined in claim 1. In a radio communication network where the radio resources of the network are allocated to a plurality of users and where a user is allocated a certain transmission capacity, the radio resources are, according to the invention, allocated to a user depending on an utilization factor which is determined relating to the transmission capacity allocated to a user. In other words, it is determined, how much of the transmission capacity, which is assigned to a particular user, is actually used by this particular user. If the actually used transmission capacity is too low compared to the assigned capacity, then the algorithm for radio resource allocation can be modified or adjusted in order to achieve a higher utilization factor.

The utilization factor is some kind of measure for the amount of radio resources, which are assigned to but not used by the user, in other words, a measure for the wasted amount of radio resources. The utilization factor can, for example, be expressed as a ratio of the used to the unused resources, as an absolute value of the unused resources or as any other suitable value.

By determining the utilization factor relating to the transmission capacity allocated to a user and taking this utilization factor into account in the radio resource allocation algorithm, the invention enables a more efficient use of the available radio resources. This in turn enables the network to serve more subscribers simultaneously.

- 5 There are different possibilities to determine the utilization factor. It could for example be determined by monitoring the absolute amount of transmitted data or by determining the existence of an active connection to another terminal. But these possibilities could lead to distorted values, because the user has transmitted a large amount of data, but could have transmitted an even larger amount or because there exists an active connection, but actually no data is transmitted.
- 10

In a preferred embodiment of the invention, the utilization factor is determined by detecting time intervals in which the user does not or not completely exploit the transmission capacity which is allocated to him. He either transmits no or not as much data as the assigned transmission capacity would allow. This enables a precise measurement how much of the transmission capacity allocated to the user is actually used by the user to transmit data.

An advantageous way to determine the utilization factor is to detect just those time intervals, in which the user does not make use of the assigned transmission capacity, that means in which the user does not transmit or receive any data at all.

- 15
- 20 The detection of these time intervals can be done in several, favoured ways. The first of them is to directly monitor the radio interface of the radio communication network and detect those time periods without any data throughput. In communication systems with a plurality of communication channels, only those channels which are allocated to the particular user have to be monitored.
- 25 This method has the advantage, that the equipment to detect the time intervals, i.e. a radio receiver, can be placed anywhere between the transmitter and the receiver, where the sig-

nal from the transmitter can be received. It furthermore permits to consider properties of the air interface when determining the utilization factor.

In a radio network, data transmission between a transmitter and a transceiver is typically done with a multilayer protocol stack where each layer is designed to carry out a specific task. Many of the current protocol stacks comply with the generally known OSI (Open Systems Interconnection) model, which specifies seven different layers. When directly monitoring the air interface, the measuring of the data rate, the number of bytes transmitted and the actual transmission time can, for example, be done by evaluating the protocols used for the transmission in the radio network.

10 In the second method it is not the air interface that is monitored, but it is the first layer of the protocol stack, which is monitored directly in the transmitter and/or the receiver. The first layer is usually known as the physical layer and performs tasks in connection with the actual bit transmission.

An advantage of this method is, that no additional hardware is necessary, because the monitoring can for example be implemented as software running on the hardware of the transmitter and/or the receiver.

It is more difficult and more demanding to determine the utilization factor if the user does transmit data, but does not completely exploit the transmission capacity which is allocated to him. So a third method can be applied in systems, where the user is allocated radio resources by allocating a data transmission rate or where the allocated radio resources can be expressed as a data transmission rate. Here, an actual transmission time is determined by measuring, how much time is needed by the user to transmit a given amount of data. Additionally a target transmission time, which relates to the shortest possible time period in which the given amount of data can be transmitted with the given data rate, is calculated by dividing the given amount of data by the data transmission rate. Next, the time intervals without data transmission are calculated by subtracting the target transmission time from the actual transmission time. Finally, the desired utilization factor can be determined.

This method also can be easily implemented in existing systems for instance as pure software to be integrated in node of a telephone network. However, to employ this method, the data transmission rate (e.g. a given number of bytes per second) as well as the amount of transmitted data (e.g. the number of bytes transmitted) have to be known or at least measurable.

The measuring of the data rate, the number of bytes transmitted and the actual transmission time can, for example, again be done by monitoring the protocols used for the transmission in the radio network.

In some communication systems, a user can be assigned more than one transmission channel simultaneously, i.e. two or more timeslots in a TDMA system or two different carrier frequencies in a FDMA system. It would be possible to determine the utilization factor relating to the joint transmission capacity of all or at least two transmission channels. However, in such systems, it could be possible, that the user continuously transmits data in one channel but only intermittently in another channel. Hence, there are time intervals with a higher throughput than other time intervals but there are no time intervals with no data throughput at all. Thus the determination of the utilization factor by detection of time intervals without any data transmission could produce incorrect results.

In order to obtain accurate results with any of the above-mentioned methods, the utilization factor is preferably determined for each transmission channel separately by separately monitoring each channel.

It is a further object of the invention to provide a radio network and a device for an efficient allocation of radio resources to a plurality of users. This object is achieved by the radio network according to claim 8 and the device according to claim 12 respectively.

A radio communication network with means for an allocation of radio resources to a plurality of users, where a user is allocated a certain transmission capacity, further includes means for the determination of the utilization factor relating to the transmission capacity allocated to a specific user. According to the invention, the means for allocating the radio

resources are formed in such a way that the radio resources are allocated to the users depending on the determined utilization factor.

Regarding the object of the invention to provide a device for an efficient allocation of radio resources of a radio communication network, it is achieved by including the means for the 5 determination of the utilization factor in this device.

From the following detailed description and from the entirety of the claims it will be clear to a person skilled in the art, that there are more advantageous embodiments and feature combinations of the invention.

Short description of the drawings

10 The drawings used for illustration of the examples show:

Fig. 1 A diagram showing the waiting time for a user and the wastage of radio resources for a particular user transmission behaviour and a specific radio resource allocation method within a radio communication network;

15 Fig. 2 another diagram showing the waiting time and the wastage of radio resources for another transmission behaviour and another allocation method;

Fig. 3 a schematical view of a part of a radio communication network according to the invention and

Fig. 4. a more detailed view of some elements of the radio network as shown in fig. 3.

20 In general, the same objects in different drawings are given the same reference numerals.

Ways of carrying out the invention

There exist different methods within a radio communication network to allocate radio resources to the users. In order to minimize the waiting time for a user, a first method allocates a particular amount of data for a long time, as soon as a user wants to transmit data.

5 This method has a disadvantage with respect to the exploitation of the totally available radio resources. This is particularly true in those cases, where the user is not able to produce enough data to send it continuously, because for example his production rate is smaller than the data transmission rate or because he does not want to send anything. The waiting time for a user and the wastage of the radio resources for such a scenario are shown in the
10 diagram of figure 1.

Time is shown as the horizontal axis of the diagram. The first row 1.1 shows the intention of the user to transmit data. Row 1.1 is alternating between high and low, because, as explained above, the user is not able to send data continuously. Row 1.2 shows the radio resources which are allocated to the user. After the first send request of the user, he is allocated a particular resource for a long time without any interruptions. Row 1.3 shows the time periods, when the user is actually transmitting data. Row 1.3 has substantially the same shape as row 1.1, which indicates that the user can send data as soon as he has some data to send. He does not have to wait until he is allocated some resources to transmit his data. The waiting time for the user is shown in row 1.4. Except a small period of time after
15 the send request, when the user has to wait until he is allocated some resources, row 1.4 is always low. Finally, row 1.5 shows the wastage of the radio resources allocated to the user. During those periods of time where the line in row 1.5 is high, radio resources are allocated to the user, but not used by him, because he is not able or does not want to transmit data. As can be seen, the percentage of the line being high is relatively high and therefore
20 a lot of radio resources are wasted. On the other hand, the user hardly has to wait to transmit data, resulting in a high data throughput for this user.
25

In the diagram of figure 2, waiting time and resource wasting are shown for a second scenario. In order to optimize the utilization of the available radio resources, a user is alloca-

replacement sheet 9

ART 34 AMDT

ted a specific radio resource only for small periods of time, no matter whether the user has a small or a large amount of data to send.

In our example in figure 2, the user has a lot of data to send, which is indicated by row 2.1, which is high almost all of the time. As shown by the alternating row 2.2, the user is allocated a specific radio resource only for small periods of time. Later on, when the resource allocation algorithm of the radio network realises that the user has a large amount of data to send, the user is allocated the resource a long period of time. In this case, row 2.3 which shows the actual transmission time, has substantially the same shape as row 2.2, which indicates that the actual sending time is not limited by the ability of the user to produce the data but by the short resource allocation periods. Accordingly, the data throughput is lowered, because the user has to wait for data transmission almost all of the time, when he is not allocated any resources. This is shown in row 2.4. The last row 2.5 again shows the wastage of the radio resources. Here the advantage of this allocation method can be seen. Only a small amount of the allocated radio resources are wasted, because the periods of time, where the user is allocated the radio resources, but does not actually transmit any data, are very small. Hence the available radio resources can be shared with other users.

To find a better radio resource allocation algorithm, for instance an algorithm allowing to find a compromise between data throughput and resource wastage, it is essential to know, how much of the radio resources allocated to a particular user are actually utilized by the user to transmit data.

According to the invention, a utilization factor, which is a measure for the amount of radio resources wasted by a user, is determined. As an example of a radio communication network, figure 3 shows a part of a mobile telephone network with a mobile switching centre (MSC) 3, two basestations 4 and 5, which are connected to the MSC 3 by communication links 6 and 7, and two mobile user terminals 8 and 9, which are connected to the basestations 4 and 5 respectively by radio links 10 and 11.

Further, MSC 3 is connected to a telephone network 14 by a communication link 15 and a second MSC 12 is connected to MSC 3 by a communication link 13. MSC 12 can serve other basestations and users (not shown) in a different geographical area.

While in a mobile telephone network the connection between the mobile terminals 8 and 9

5 and the basestations 4 and 5 are always radio links, the communication links 6, 7 between the basestations 4,5 and the MSC 3 as well as the communication links 13, 15 from or to the MSC 3 can be of a any type, including for example cable, radio or glass fibre links.

Usually, a mobile telephone network comprises further components, which are not essential to the invention and therefore not shown in the drawings.

10 As an example, we assume that the utilization factor has to be determined for the user of terminal 8, called user A; when, for instance, communicating with the user of terminal 9, called user B. This monitoring of the communication activities of user A can be done in different locations. In figure 3, monitoring devices 16.1 – 16.5 are shown in some of these locations.

15 Monitoring device 16.1 is located within the mobile switching centre MSC 3 and is built for monitoring the communication channels used by user A. Monitoring device 16.1 could directly monitor the data packets sent from or to user A on the incoming or outgoing links of MSC 3. In this case, it would have to evaluate the data packets by itself. Another possibility would be to benefit from the fact, that the MSC 3 does evaluate the data packets anyway and therefore use these results to monitor the transmission protocols and to determine the utilization factor.

20

Monitoring device 16.2 directly monitors the communication link 6 between MSC 3 and basestation 4 by tapping this link. How this link is tapped depends on the physical type (radio link, cable etc.) of link.

25 Monitoring device 16.3 is located within basestation 4. The monitoring can be done in the same ways as done by monitoring device 16.1.

Monitoring device 16.4, which monitors directly the air interface by receiving and evaluating the transmission signals sent by terminal 8 or the basestation 4, is located somewhere within the radio cell served by basestation 4, where it can receive the signals from the basestation 4 as well as those from terminal 8.

5 Monitoring device 16.5 is directly connected to terminal 8. It can evaluate the data packets by itself or, similar to monitoring device 16.1 can benefit from the data packet evaluation of the terminal 8.

If user A communicates with a user connected to MSC 12 or with a user within the telephone network 14, the controlling of his communication could also be done with a monitoring device connected to the communication links 13 or 15 or even with a monitoring device somewhere within the telephone network 14.

10 Each of the monitoring devices 16.1 – 16.5 can, as shown in figure 3, be implemented as an independent apparatus in a separate housing which then has to be connected to the other components of the network as required. However, the monitoring devices 16.1 – 15 16.5 can also be integrated in other components of the network by means of hardware and/or software.

In order to determine a utilization factor to control the radio resource allocation method, one of the shown monitoring devices 16.1 – 16.5 is sufficient. Nevertheless, two or more monitoring devices 16.1 – 16.5 could be used simultaneously to determine two or more utilization factors. Here the resource allocation could depend on some or all of these utilization factors.

20 Figure 4 shows a more detailed view of the monitoring device 16.4 and MSC 3, which, according to our example, allocates the required radio resources to the terminals 8 and 9 with an allocation unit 21. The resource allocation is done depending on this utilization factor, that is why it has to be communicated to the MSC 3. To transmit the utilization factor to MSC 3, any of the existing or any other suited communication link could be used.

The monitoring of the communication between basestation 4 and terminal 8 by the monitoring device 16.4 is done by receiving the transmission signals from terminal 8 and/or basestation 4, transmitted on radio link 10. Next, the monitoring device 16.4 detects with a detector 18 time periods in the received signals where the terminal 8 does neither send 5 nor receive any data. A processor unit 19 then determines a utilization factor and transmits this factor to the MSC 3. For this it utilizes the communication link 20, which is shown as a dotted line, because the type of communication link depends on where the monitoring device 16.4 is located. In our example, the utilization factor would probably be transmitted to the MSC 3 by a radio link from the monitoring device 16.4 to the basestation 4 and further from basestation 4 to MSC 3 via communication link 6.

The utilization factor as received by the MSC 3 is forwarded to the allocation unit, which, depending on the value of the utilization factor, adjusts the currently used resource allocation algorithm in order to obtain a better resource exploitation by the terminals.

To summarise it can be stated that the invention enables the improvement of the algorithms, which are used to allocate the available radio resources in a radio communication 15 network to the users, by measuring how much of the radio resources allocated to a user are actually used, in other words, how much of these resources are wasted. By repeating this measurements, the end result should be a radio network, where the transmission capacity is used efficiently to serve the maximum number of subscribers simultaneously.